Content-based Video Indexing for the Support of Digital Library Search

M. Petković, R. van Zwol, H.E. Blok, W. Jonker, P.M.G. Apers
University of Twente
Enschede, The Netherlands
{milan,zwol,blok,jonker,apers}@cs.utwente.nl

M. Windhouwer, M. Kersten
CWI
Amsterdam, The Netherlands
{windhouw,mk}@cwi.nl

1. Introduction

Recent advances in computing, communications and data storage have led to an increasing number of large digital libraries, which are nowadays publicly available on the Internet. However, to find required information in that enormous mass of data becomes very difficult. In this demo we present a digital library search engine that combines efforts of the AMIS and DMW research projects, each covering significant parts of this problem.

The most important contributions of our work are the following:

1. We demonstrate a flexible solution for extraction and querying of meta-data from multimedia documents in general [3].
2. Scalability and efficiency support are illustrated for full text indexing and retrieval [1].
3. We show how for a more limited domain, like an Intranet, conceptual modeling can offer additional and more powerful query facilities [4].
4. In the limited domain case, we demonstrate how domain knowledge can be used to interpret low-level features into semantic content [2].

In this short description, we will focus on the first and the fourth item. For the other two items, as well as for a more detailed description of our architecture of the search engine we refer to the demo web site:
`http://www.cs.utwente.nl/~dmw/ICDE2002/`

2. Motivating example

In order to demonstrate advantages of the proposed system architecture, we constructed a specialized search engine for the Australian Open tennis tournament web site (see `http://tournament.ausopen.org/`). This web site is a good example of a site with a hidden semantical structure. Some semantic concepts, which were clearly available in the source data used for this page, are lost due to the translation of the source data into HTML. As a result, search engines can only see the page as a body of text and provide the ability to search for keywords in it. Alternatively, using the webspace method [4] and conceptual information, the query could be formulated more precise.

Apart from conceptual information, the site also contains multimedia fragments, like audio files of interviews and even videos of tennis matches. So, content-based retrieval of the hidden information inside the multimedia content can also be offered to the user. This enables them to issue queries which combine concept and content-based information such as: *"Show me video scenes of left-handed female players who have won the Australian Open in the past, in which they approach the net."*

In the next section, we will focus on extraction of this hidden information inside the multimedia content in our specific domain, but also we will briefly describe a technique that can manage extraction of such multimedia meta data in any domain.

3. Video indexing

In order to explore video content and provide a framework for automatic extraction of semantic content from raw video data, we propose the *COntent-Based RetrievAl* (COBRA) video data model [2]. The model is independent of feature/semantic extractors, providing flexibility by using different video processing and pattern recognition techniques for that purposes. At the same time it is in line with the latest development in MPEG-7, distinguishing four distinct layers within video content: the raw data, the feature, the object, and the event layer. The object and event layers consist of entities characterized by prominent spatial and temporal dimensions respectively.

As a video is a temporal sequence of pixel regions at the physical level, it is very difficult to explore its semantic content, i.e. detect and recognize video events automatically. Very often, different models and techniques are required for event understanding. Furthermore, as it is shown in the lit-

Proceedings of the 18th International Conference on Data Engineering (ICDE'02)
1063-6382/02 $17.00 © 2002 IEEE

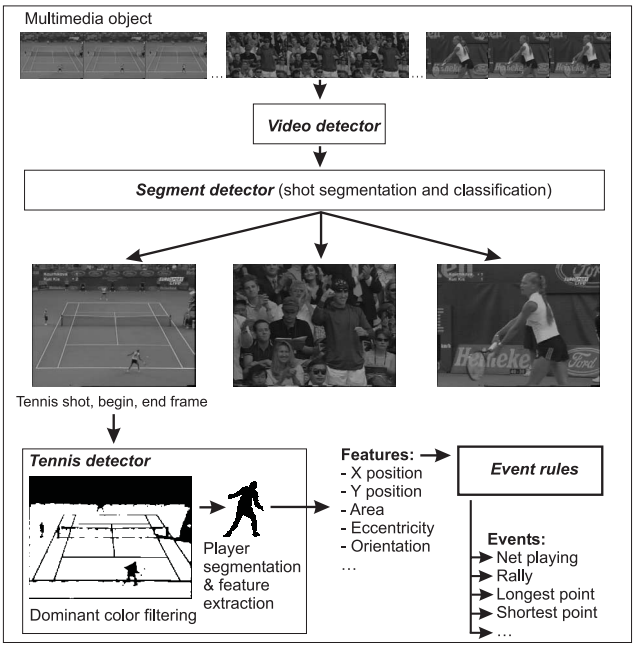

Figure 1. Tennis FDE: detector dependencies.

erature, different techniques are more suitable for different kinds of events.

To provide automatic extraction of concepts (objects and events), the COBRA video model is enriched with a few extensions. Each extension encapsulates a different knowledge-based technique for concept inference.

First, the model is extended with object and event grammars. These grammars are aimed at formalizing the descriptions of high-level concepts, as well as facilitating their extraction based on spatio-temporal reasoning. The syntax of these rules is described in [2], where they were implemented within the query engine.

We formed an instantiation of the COBRA framework for our specific domain using a feature grammar. The feature grammar, which forms the core of the Acoi system [3], describes the relationships between meta-data and detectors in a set of grammar rules. A tennis feature grammar with rules that describe the execution order of and dependecies between several feature, object or event extraction algorithms has been develeoped (see Figure 1). Managing the meta-index now boils down to exploiting the dependencies in the feature grammar. For example, to populate the meta-index the feature grammar is used to generate a parser: the *Feature Detector Engine* (FDE). This FDE triggers the execution of the associated detectors. In the remainder we will describe these detection steps.

After finding a video object, the tennis FDE executes the segment detector. This detector, which is implemented externally, segments the video into different shots. The shot boundaries are detected using differences in color histograms of neighboring frames. The same algorithm encapsulates shot classification. It classifies shots in four different categories: tennis, close-up, audience, and other. The court shots are recognized based on the dominant color. A shot is classified as close-up, if it contains a significant amount of skin colored pixels. For the classification, we also use entropy characteristics, mean and variance.

If the segment detector classifies a shot as a tennis shot, the tennis FDE starts execution of the tennis detector. Its task is to segment and track the tennis player. Using estimated statistics of the tennis field color, the algorithm does the initial quadratic segmentation of the first image of a video sequence classified as a playing shot. In the next frames, we predict the player position and search for a similar region in the neighborhood of the initially detected player.

In the same algorithm we extract features characterizing the shape of the segmented player's binary representation. Having the specific case of the human figure in this particular application, we extract special parameters trying to maximize their informative value. Besides the player's position, we extract the dominant color, and standard shape features such as the mass center, the area, the bounding box, the orientation, and the eccentricity.

Player's positions and their transitions over time are related to particular events (net-playing, rally, etc.) using rules. These rules, which use spatio-temporal relations, are implemented as white- and blackbox detectors within the FDE.

All this accumulated video meta-data can now be used to support the content-based parts, e.g. video scenes showing a net-play event, of digital library queries.

References

[1] H. E. Blok, A. P. d. Vries, H. M. Blanken, and P. M. Apers. Experiences with IR TOP N Optimization in a Main Memory DBMS: Applying 'the Database Approach' in New Domains. In *Advances in databases, 18th British National Conference on Databases, BNCOD 18, Lecture Notes in Computer Science, Springer*, 2001.

[2] M. Petkovic and W. Jonker. Content-based video retrieval by integrating spatio-temporal and stochastic recognition of events. In *proceedings of IEEE Intl. Workshop on Detection and Recognition of Events in Video*, Vancouver, Canada, 2001.

[3] M. A. Windhouwer, A. R. Schmidt, and M. L. Kersten. Acoi: A System for Indexing Multimedia Objects. In *proceedings of International Workshop on Information Integration and Web-based Applications & Services*, Yogyakarta, Indonesia, November 1999.

[4] R. v. Zwol and P. Apers. The webspace method: On the integration of database technology with information retrieval. In *proceedings of CIKM'00*, Washington, DC., Nov. 2000.